On the robustness of LISREL (maximum likelihood estimation) against small sample size and non-normality.

Boomsma, Anne

2. *ANALYSIS OF STRUCTURAL EQUATION MODELS*

In this chapter the general structural model called LISREL will be described in mathematical detail, dealing with the main objectives of the use of such models, model fitting and estimation of parameters. The reasons why we choose to investigate the robustness against small sample size and non-normality are discussed in section 2.4. First a small amount of background information is given.

2.1 Some references

In planning this book we intended to include a brief historical account of structural equation models. Given recent reviews in the literature (e.g. Bielby & Hauser, 1977; Bentler, 1980; Saris, 1979a, p. 32 ff., 1980) this is no longer necessary. The papers of Bielby & Hauser and Bentler with hundreds of references give also full attention to the main fields of application of structural models: sociology, econometrics, educational and psychological sciences. It can be noticed from their reviews that in the late seventies little was known about the robustness of maximum likelihood estimators against violation of assumptions in LISREL-type models (see Bielby & Hauser, 1977, p. 153 and Bentler, 1980, p. 444 ff.).

Before the LISREL-model is defined it must be stressed that this model can be viewed as a special case of a more general approach to the analysis of covariance structures (see Jöreskog, 1981). The latter approach is more general first, because it has no restrictions with respect to the form of $\underset{\sim}{\Sigma}$, so that any covariance structure can be handled, secondly, three different methods of estimation (unweighted least squares, generalized least squares and maximum likelihood) can be used. The LISREL-model on the other hand assumes a definite form of $\underset{\sim}{\Sigma}$ and, until the LISREL-V version was made public, maximum likelihood estimation. According to Jöreskog (o.c., p. 76) the LISREL-model is still very general and in contrast to the more general approach so flexible, that it can handle almost any structural equation problem arising in practice.

Less general models for the analysis of covariance structures have been introduced earlier, for example by Bock (1960) and Bock & Bargmann (1966); see also Wiley, Schmidt & Bramble (1973) who present a generalization of the Bock & Bargmann class of models.

After Jöreskog's introduction of the LISREL-model several other models have been discussed in the literature, which are either viewed as generalizations, simplifications, or different representations of the LISREL-model. See for example Bentler (1976), Weeks (1978, 1980), Bentler & Weeks (1979, 1980), who define models for the analysis of moment structures; McDonald (1978, 1979) with the COSAN-program for a class of general models; and McArdle (1979, 1980), Horn & McArdle (1980), McArdle & McDonald (1980) with the RAM-representation. All these different models for the analysis of covariance and moment structures are not discussed here. The only thing to be noted is that from a very general, but impractical and inflexible class of models, the LISREL-model, with its specific covariance structure and its maximum likelihood estimation method, is chosen to be studied for its robustness characteristics. There is one simple reason why this particular model was chosen: when we started to think about this dissertation it was the most general model with an available computer program, ready for use. It is very unlikely that at this very moment, 1983, our choice would be a different one, because the prevailing practical use of the LISREL-program makes answers as to its robustness most urgent.

Finally, some attention should be given to a type of modeling known as "soft modeling", or as the PLS (Partial Least Squares)-approach. This general, distribution-free method for path models with latent variables has been developed chiefly by Wold (1978, 1980, 1982). He frequently states that PLS modeling is primarily designed for causal predictive analysis of problems with high complexity and low information, and emphasizes that the maximum likelihood LISREL and partial least squares approaches to path models with latent variables are complementary rather than competitive. A more detailed comparison of maximum likelihood and PLS techniques is given by Dijkstra (1981) and by Jöreskog & Wold (1982).

2.2 *The general LISREL-model*

For the general model a notation was chosen which is almost identical to the notation used by Jöreskog (e.g. 1977) and to the one found in program descriptions of LISREL (e.g. Jöreskog & Sörbom, 1978). Two deviations should be noted: N instead of M will be used for the number of independent observations, and $k = p + q$ will symbolize the total number of observed random variables.

The general model specifies a linear structural relationship between a random vector of *latent dependent* (endogenous) variables $\underset{\sim}{\eta} = (\eta_1,\eta_2,\ldots,\eta_m)'$ and a random vector of *latent independent* (exogenous) variables $\underset{\sim}{\xi} = (\xi_1,\xi_2,\ldots,\xi_n)'$, which relation is expressed by

$$\underset{\sim}{B}\underset{\sim}{\eta} = \underset{\sim}{\Gamma}\underset{\sim}{\xi} + \underset{\sim}{\zeta}, \tag{2.1}$$

where $\underset{\sim}{B}$ (m x m) and $\underset{\sim}{\Gamma}$ (m x n) are coefficient matrices and $\underset{\sim}{\zeta} = (\zeta_1,\zeta_2,\ldots,\zeta_m)'$ is a random vector of *errors in equations* (residuals, disturbances). It is assumed that $\underset{\sim}{B}$ is non-singular and that $\underset{\sim}{\xi}$ is uncorrelated with $\underset{\sim}{\zeta}$.

Corresponding to the unobserved vectors of latent variables $\underset{\sim}{\eta}$ and $\underset{\sim}{\xi}$ there are random vectors of *observed* variables $\underset{\sim}{y} = (y_1,y_2,\ldots,y_p)'$ and $\underset{\sim}{x} = (x_1,x_2,\ldots,x_q)'$. The relation between latent variables and observed variables is expressed by

$$\underset{\sim}{y} = \underset{\sim}{\mu} + \underset{\sim}{\Lambda}_y\underset{\sim}{\eta} + \underset{\sim}{\varepsilon}, \tag{2.2}$$

$$\underset{\sim}{x} = \underset{\sim}{\nu} + \underset{\sim}{\Lambda}_x\underset{\sim}{\xi} + \underset{\sim}{\delta}, \tag{2.3}$$

where $\underset{\sim}{\Lambda}_y$(p x m) and $\underset{\sim}{\Lambda}_x$(q x n) are regression matrices of $\underset{\sim}{y}$ on $\underset{\sim}{\eta}$ and of $\underset{\sim}{x}$ on $\underset{\sim}{\xi}$, respectively, with $\underset{\sim}{\varepsilon}$ and $\underset{\sim}{\delta}$ as the corresponding vectors of errors in the observed variables (*errors of measurement*), while $\underset{\sim}{\mu} = E(\underset{\sim}{y})$ and $\underset{\sim}{\nu} = E(\underset{\sim}{x})$ by definition.

It is assumed that the errors of measurement are uncorrelated with $\underset{\sim}{\eta}$, $\underset{\sim}{\xi}$ and $\underset{\sim}{\zeta}$, but they may be correlated among themselves. Without loss of generality it is furthermore assumed that $E(\underset{\sim}{\eta}) = E(\underset{\sim}{\zeta}) = \underset{\sim}{0}$ and $E(\underset{\sim}{\xi}) = \underset{\sim}{0}$.[As shown by Sörbom (1982), these assumptions are not necessary and the LISREL-program can also estimate the means $E(\underset{\sim}{\xi})$ and $E(\underset{\sim}{\eta})$.]

It follows that $E(\underset{\sim}{\varepsilon})=\underset{\sim}{0}$ and $E(\underset{\sim}{\delta})=\underset{\sim}{0}$.

Finally, the covariance matrices of $\underset{\sim}{\xi}$, $\underset{\sim}{\zeta}$, $\underset{\sim}{\varepsilon}$ and $\underset{\sim}{\delta}$ are denoted by $\underset{\sim}{\Phi}$ (n x n), $\underset{\sim}{\Psi}$ (m x m), $\underset{\sim}{\Theta}_{\varepsilon}$(p x p) and $\underset{\sim}{\Theta}_{\delta}$(q x q), respectively.

Let $\underset{\sim}{z} = (\underset{\sim}{y}',\underset{\sim}{x}')'$ be the vector of k = p + q observed random variables. The covariance matrix $\underset{\sim}{\Sigma}$ (k x k) of $\underset{\sim}{z}$ can then be derived from the above assumptions (e.g. Dijkhuizen, 1978, p.7):

$$\underset{\sim}{\Sigma} = \begin{bmatrix} \underset{\sim}{\Lambda}_y (\underset{\sim}{B}^{-1}\underset{\sim}{\Gamma}\underset{\sim}{\Phi}\underset{\sim}{\Gamma}'\underset{\sim}{B}'^{-1} + \underset{\sim}{B}^{-1}\underset{\sim}{\Psi}\underset{\sim}{B}'^{-1})\underset{\sim}{\Lambda}'_y + \underset{\sim}{\Theta}_{\varepsilon} & \underset{\sim}{\Lambda}_y\underset{\sim}{B}^{-1}\underset{\sim}{\Gamma}\underset{\sim}{\Phi}\underset{\sim}{\Lambda}'_x \\ \underset{\sim}{\Lambda}_x\underset{\sim}{\Phi}\underset{\sim}{\Gamma}'\underset{\sim}{B}'^{-1}\underset{\sim}{\Lambda}'_y & \underset{\sim}{\Lambda}_x\underset{\sim}{\Phi}\underset{\sim}{\Lambda}'_x + \underset{\sim}{\Theta}_{\delta} \end{bmatrix} . \qquad (2.4)$$

It is easily seen that the elements of $\underset{\sim}{\Sigma}$ are functions of the elements of $\underset{\sim}{\Lambda}_y$, $\underset{\sim}{\Lambda}_x$, $\underset{\sim}{B}$, $\underset{\sim}{\Gamma}$, $\underset{\sim}{\Phi}$, $\underset{\sim}{\Psi}$, $\underset{\sim}{\Theta}_{\varepsilon}$ and $\underset{\sim}{\Theta}_{\delta}$. The elements of these eight matrices are of three kinds: a) *fixed* parameters, having assigned values, b) *constrained* parameters, unknown but equal to at least one other parameter, c) *free* parameters, unknown and not constrained to be equal to any other parameter. The vector of all independent constrained (counting each distinct constrained parameter once only) and free parameters will be denoted by $\underset{\sim}{\omega} = (\omega_1,\omega_2,\ldots,\omega_s)'$. In section 4.2 (page 51) an example is given of a model with a specific covariance structure $\underset{\sim}{\Sigma}$. In that particular example it is shown of what type the elements of the eight matrices are, they are either fixed or free. Finally, it is illustrated there how the free parameters (the elements of $\underset{\sim}{\Sigma}$ which have to be estimated given a sample of observations) can be put one by one in a vector $\underset{\sim}{\omega}$.

The general model is defined by equations (2.1), (2.2) and (2.3). More specifically, the *structural equation model* is constituted by equation (2.1), and the *measurement model* by equations (2.2) and (2.3). It should be mentioned that there is a distinction between the *model* as defined by the three equations and the *program* used by the researcher in order to get estimates of the unknown parameters of the model. In this study we use the term LISREL both for the model and for the program. It will be clear from the context in which we use the term, whether the model or the program is meant. Others (e.g. Bentler & Weeks, 1980, p. 290) would strictly insist on a distinction between what is called the Jöreskog-Keesling-Wiley model (cf. Keesling, 1972; Wiley, 1973) and the Jöreskog-Sörbom program.

So far, we have only described the model in terms of the *population*. The general aim of the use of LISREL is to estimate the s model parameters, given the observed *sample* values on k variables for N individuals (or other objects under study). It is from here where the statistical assumptions start to play a role. Before the essentials of structural parameter estimation by LISREL are summarized it should be noted that only in this chapter a distinction is made between the stochastic values of the variables (denoted by a vector $\underset{\sim}{z}$ of length k) and the realized values of those random variables (denoted by $\underset{\sim}{\mathcal{z}}$).

Given N independent identically distributed observations $\underset{\sim}{\mathcal{z}}_1, \underset{\sim}{\mathcal{z}}_2, \ldots, \underset{\sim}{\mathcal{z}}_N$ of $\underset{\sim}{z} = (\underset{\sim}{y}', \underset{\sim}{x}')'$ with expectation $(\underset{\sim}{\mu}', \underset{\sim}{\nu}')'$ and covariance matrix $\underset{\sim}{\Sigma}$ as defined by (2.4), estimates of the unknown parameters $\underset{\sim}{\omega}$ in $\underset{\sim}{\Lambda}_y$, $\underset{\sim}{\Lambda}_x$, $\underset{\sim}{B}$, $\underset{\sim}{\Gamma}$, $\underset{\sim}{\Phi}$, $\underset{\sim}{\Psi}$, $\underset{\sim}{\Theta}_\varepsilon$ and $\underset{\sim}{\Theta}_\delta$ are required. Assuming that the distribution of $\underset{\sim}{z}$ is multivariate normal (N_k) it is possible to get *maximum likelihood estimates* of the elements of $\underset{\sim}{\omega}$, having nice distributional properties. The estimates derived with LISREL are based on the sample covariance matrix $\underset{\sim}{S}$ (k x k). A more detailed presentation of the properties of the estimates follows in the next section .

2.3 *Estimation of parameters*

Let $\underset{\sim}{Z} = (\underset{\sim}{\mathcal{z}}_1, \underset{\sim}{\mathcal{z}}_2, \ldots, \underset{\sim}{\mathcal{z}}_N)'$, where $\underset{\sim}{\mathcal{z}}_i$ is a vector of length k, denote the N independently realized values of random variables $\underset{\sim}{z} = (\underset{\sim}{y}', \underset{\sim}{x}')'$. And let $f(\underset{\sim}{Z}; \underset{\sim}{\omega})$ denote the joint probability density function of $\underset{\sim}{Z}$, where $\underset{\sim}{\omega} = (\omega_1, \omega_2, \ldots, \omega_s)'$ is a vector of $s \leq \frac{1}{2}k\ (k + 1)$ unknown parameters, which belongs to the set of permissible values for $\underset{\sim}{\omega}$. This set is called the parameter space and denoted by $\underset{\sim}{\Omega}$. The likelihood of $\underset{\sim}{\omega} \in \underset{\sim}{\Omega}$, given the observations, is a function of $\underset{\sim}{\omega}$:

$$L(\underset{\sim}{\omega}; \underset{\sim}{Z}) = f(\underset{\sim}{Z}; \underset{\sim}{\omega}).$$

In practice, it is convenient to work with the natural logarithm of the likelihood, denoted by $l(\underset{\sim}{\omega};\underset{\sim}{Z}) = \log f(\underset{\sim}{Z};\underset{\sim}{\omega})$.

The principle of maximum likelihood consists of taking that $\hat{\underset{\sim}{\omega}} = (\hat{\omega}_1,\hat{\omega}_2,\ldots,\hat{\omega}_s)'$ as the estimate of $\underset{\sim}{\omega}$, for which

$$l(\hat{\underset{\sim}{\omega}};\underset{\sim}{Z}) = \sup_{\underset{\sim}{\omega}\in\underset{\sim}{\Omega}} l(\underset{\sim}{\omega};\underset{\sim}{Z}) .$$

In our case $\underset{\sim}{z}$ has a multivariate normal distribution with $E(\underset{\sim}{z}) = (\underset{\sim}{\mu}',\underset{\sim}{\nu}')'$ and covariance matrix $\underset{\sim}{\Sigma}$. The sample covariance matrix of $\underset{\sim}{z}$ is defined by

$$\underset{\sim}{S} = (N-1)^{-1} \sum_{i=1}^{N} (\underset{\sim}{z}_i-\bar{\underset{\sim}{z}})(\underset{\sim}{z}_i-\bar{\underset{\sim}{z}})' , \tag{2.5}$$

where $\bar{\underset{\sim}{z}} = N^{-1} \sum_{i=1}^{N} \underset{\sim}{z}_i$. This sample covariance matrix is an unbiased estimate of $\underset{\sim}{\Sigma}$. It can be proved (Anderson, 1958, p. 159) that the distribution of $\underset{\sim}{S}$ is a Wishart distribution with parameter matrix $(N-1)^{-1}\underset{\sim}{\Sigma}$ and $N-1$ degrees of freedom, shortly $W_k[(N-1)^{-1}\underset{\sim}{\Sigma}, N-1]$.

Because the first moment in the population is unconstrained, the log likelihood function $l(\underset{\sim}{\Sigma};\underset{\sim}{S})$ can be used directly instead of considering the log likelihood function $l(\underset{\sim}{\mu},\underset{\sim}{\nu},\underset{\sim}{\Sigma};\bar{\underset{\sim}{x}},\bar{\underset{\sim}{y}},\underset{\sim}{S})$, which is equal to $\log f(\underset{\sim}{Z};\underset{\sim}{\mu},\underset{\sim}{\nu},\underset{\sim}{\Sigma})$; as noted before, here we have no interest in getting estimates of $\underset{\sim}{\mu}$ and $\underset{\sim}{\nu}$ (cf. Sörbom, 1974, 1976, 1982, for examples where such an interest does exist; cf. also Lee & Tsui, 1982). Now it can be derived that the likelihood function of $\underset{\sim}{\Sigma}$, given $\underset{\sim}{S}$, is

$$L(\underset{\sim}{\Sigma};\underset{\sim}{S}) = f(\underset{\sim}{S};\underset{\sim}{\Sigma}) = c|\underset{\sim}{\Sigma}|^{-\frac{1}{2}(N-1)} \exp\,[-\tfrac{1}{2}(N-1)\ \mathrm{tr}\ \underset{\sim}{S}\underset{\sim}{\Sigma}^{-1}], \tag{2.6}$$

where c is a suitable constant (Anderson, 1958, p. 157). The corresponding log likelihood function

$$l(\underset{\sim}{\Sigma};\underset{\sim}{S}) = \log f(\underset{\sim}{S};\underset{\sim}{\Sigma}) = -\tfrac{1}{2}(N-1)\ [\log|\underset{\sim}{\Sigma}| + \mathrm{tr}\ \underset{\sim}{S}\underset{\sim}{\Sigma}^{-1}] + \log c, \tag{2.7}$$

follows from (2.6).

Since the elements of $\underset{\sim}{\Sigma}$ can be expressed as functions of $\underset{\sim}{\omega}$(page 14), $f(\underset{\sim}{S};\underset{\sim}{\Sigma}) = f(\underset{\sim}{S};\underset{\sim}{\omega})$. Thus maximum likelihood estimates for $\underset{\sim}{\omega}$ are defined by

$$l(\hat{\underset{\sim}{\omega}};\underset{\sim}{S}) = \sup_{\underset{\sim}{\omega}\in\underset{\sim}{\Omega}} l(\underset{\sim}{\omega};\underset{\sim}{S}) = \sup_{\underset{\sim}{\omega}\in\underset{\sim}{\Omega}} \log f(\underset{\sim}{S};\underset{\sim}{\omega}).$$

The maximum likelihood estimates $\hat{\underset{\sim}{\omega}}$ are found by maximizing (2.7). The estimates $\hat{\underset{\sim}{\omega}}$, which maximize the log likelihood as defined by (2.7) are the same as those which *minimize* the function

$$F^* = \log |\underset{\sim}{\Sigma}| + \operatorname{tr} \underset{\sim}{S}\,\underset{\sim}{\Sigma}^{-1} . \tag{2.8}$$

Often the following equivalent function is chosen to be minimized

$$F = \log |\underset{\sim}{\Sigma}| + \operatorname{tr} \underset{\sim}{S}\,\underset{\sim}{\Sigma}^{-1} - \log |\underset{\sim}{S}| - k . \tag{2.9}$$

It is clear that the value of F equals 0 if $\underset{\sim}{S} = \underset{\sim}{\Sigma}$.

Thus the maximum likelihood estimate of $\underset{\sim}{\omega}$ is found by accepting that $\hat{\underset{\sim}{\omega}} = (\hat{\omega}_1,\hat{\omega}_2,\ldots,\hat{\omega}_s)'$ as an estimate of $\underset{\sim}{\omega}$ for which

$$F(\hat{\underset{\sim}{\omega}};\underset{\sim}{S}) = \inf_{\underset{\sim}{\omega}\in\underset{\sim}{\Omega}} F .$$

The central part of the LISREL-program is the algorithm used for minimizing the function F. The iterative minimization method is due to Davidon (1959) and described by Fletcher & Powell (1963). The LISREL-program follows the Fletcher-Powell algorithm with a few modifications (see Jöreskog, 1969 and Gruvaeus & Jöreskog, 1970,for details).

Under certain conditions maximum likelihood estimates have desirable asymptotic properties. If $\underset{\sim}{Z} = (\underset{\sim}{z}_1,\underset{\sim}{z}_2,\ldots,\underset{\sim}{z}_N)'$ is a sample with the density function $f(\underset{\sim}{Z};\underset{\sim}{\omega})$, where certain regularity conditions are satisfied, then it can be shown (cf. Cramér, 1946, p. 500 ff; Wilks, 1962, p. 379 ff.) that the joint maximum likelihood estimators have the following properties:

(i) they are consistent, and thus tend towards unbiased estimators as N increases;

(ii) asymptotically they are efficient, attaining minimum variance for large N;

(iii) asymptotically the sampling distribution of $\hat{\underset{\sim}{\omega}}$ is multivariate normal $N_s[\underset{\sim}{\omega}, \underset{\sim}{I}_{\underset{\sim}{\omega}}^{-1}]$, where $\underset{\sim}{I}_{\underset{\sim}{\omega}}$ is the so-called information matrix (cf. Rao, 1973, p. 329 ff.).

The elements of diag $(\underset{\sim}{I}_{\underset{\sim}{\omega}}^{-1})$ are lower bounds of the corresponding variances of unbiased estimators of ω_i. It should be noted that in our case asymptotically $\hat{\underset{\sim}{\omega}}$ is an unbiased estimate of $\underset{\sim}{\omega}$. It follows that asymptotically $(\hat{\underset{\sim}{\omega}} - \underset{\sim}{\omega})$ is distributed as $N_s(\underset{\sim}{0}, \underset{\sim}{I}_{\underset{\sim}{\omega}}^{-1})$. Given $\underset{\sim}{I}_{\underset{\sim}{\omega}}^{-1}$, standard errors of $\hat{\underset{\sim}{\omega}}$ can be computed by taking the square root of the diagonal elements from $\underset{\sim}{I}_{\underset{\sim}{\omega}}^{-1}$. With a finite sample size estimates of standard errors are obtained. It is then possible to compute approximate confidence intervals for $\omega_i (i=1,2,\ldots,s)$.

The properties discussed above are also valid for restricted maximum likelihood estimators, with fixed or constrained parameters $[s < \frac{1}{2}k(k+1)]$. This occurs when additional knowledge about the true parameters is available or postulated, or when certain restrictions on the vector of parameters have to be made in order to specify a model which is identified. Under such conditions the asymptotic properties of $\hat{\underset{\sim}{\omega}}$ still hold (Silvey, 1975, p.79 ff.).

When $\underset{\sim}{z}$ has a multivariate normal distribution, which is one of the basic assumptions of LISREL, the regularity conditions referred to above are satisfied, implying that when the observed random variables have such a distribution, the asymptotic properties just mentioned hold. Although the asymptotic sampling theory for maximum likelihood estimators is limited to independent, identically distributed random variables, Hoadley (1971) established conditions under which maximum likelihood estimators are consistent and asymptotically normal in the case where the observations are independent but not identically distributed.

In concluding this section the asymptotic character of the properties of maximum likelihood estimators should be stressed above all. Therefore, the first aim of the present study is to investigate the behavior of the sampling distributions of the estimators when the finite sample is of small to moderate size. A second goal is to study the effect of departures from multivariate normality.

2.4 Tests of hypotheses

Once maximum likelihood estimates have been obtained in large samples, the goodness of fit of the specified model may be tested by the likelihood ratio technique. Under sampling and distributional conditions described in the previous section and given the specification of fixed, constrained and free parameters let H_0 be the null hypothesis that the model, thus specified, holds. Now, two general cases can be considered.

(i) Let H_1 be the alternative hypothesis that $\underset{\sim}{\Sigma}$ is *any* positive definite matrix. Then

$$-2 \log \lambda = -2 \log \frac{L(\hat{\underset{\sim}{\omega}}_0;\underset{\sim}{Z})}{L(\underset{\sim}{\Omega}\ ;\underset{\sim}{Z})} \quad , \tag{2.10}$$

where $L(\hat{\underset{\sim}{\omega}}_0;\underset{\sim}{Z})$ is the likelihood under H_0 with $r = \frac{1}{2}k(k+1)-s$ restrictions posed, s being the number of independent parameters under H_0, while $L(\underset{\sim}{\Omega};\underset{\sim}{Z})$ is the maximum of the likelihood under H_1 with no restrictions imposed. It can be shown that minus twice the logarithm of the likelihood ratio equals $(N-1)F_0$, where F_0 is the minimum value of F defined in (2.9). When H_0 holds, it can be proved (cf. Kendall & Stuart, 1973, p.240 f.; Silvey, 1975, p. 113 ff.) that asymptotically, in large samples, $(N-1)F_0$ is distributed as chi-square with $\frac{1}{2}k(k+1)-s = r$ degrees of freedom.

(ii) Let H_1 be an alternative hypothesis which is less restrictive than a specified H_0, which is a *part of* the parametric structure of the model under H_1. So H_1 is not just some specified model with less restrictions, but a closely related model. Then in large samples it is possible to test H_0 against H_1. The likelihood ratio test is now

$$-2 \log \lambda = -2 \log \frac{L(\hat{\omega}_0; Z)}{L(\hat{\omega}_1; Z)} , \quad (2.11)$$

where $L(\hat{\omega}_0; Z)$ is the likelihood under H_0 and $L(\hat{\omega}_1; Z)$ is the likelihood under H_1, $r_0 > r_1$ being the respective number of restrictions imposed, and $s_0 < s_1$ the corresponding number of independent parameters to be estimated. Under H_0 and H_1 the minimum of F as defined in (2.9) is denoted by F_0 and F_1, respectively. It follows that $F_0 \geqslant F_1$ and $-2 \log \lambda = (N-1)(F_0-F_1)$. Under H_0, $-2 \log \lambda$ is asymptotically distributed as chi-square with $s_1 - s_0$ degrees of freedom.

By this asymptotic theory the goodness of fit of different models for the same data can be compared (e.g. Jöreskog, 1974; Jöreskog & Sörbom, 1977). Although in this section much emphasis was on statistical tests for goodness of fit, it should be mentioned that in practical situations especially when the sample size is not large, the likelihood ratio statistics are used as indices of fit rather than as formal tests of goodness of fit. In an exploratory analysis looking at residuals $(S-\hat{\Sigma})$ will be an indispensable help in the search for a model that fits well. Other means for assessing the fit of the model are mentioned by Jöreskog (1981). Mellenbergh (1980) provides a general discussion on model fitting. Bentler & Bonett (1980) evaluate the goodness of fit in the analysis of covariance structures by means of coefficients which do not depend on the sample size. Eiting & Mellenbergh (1980) are working within a decision theoretic framework of hypothesis testing using Monte Carlo procedures. From such Monte Carlo studies an optimal combination of number of observations, significance level and power is determined (cf. Eiting, 1981; Kelderman, Mellenbergh & Elshout, 1981).

Finally, it is emphasized again that the results above hold only if the conditions for the asymptotic normality and asymptotic efficiency of the maximum likelihood estimators are satisfied, which will be the case if z has a multivariate normal distribution. But even if the latter condition is met, it is of interest to look at the sampling distribution of $-2 \log \lambda$ for small and moderate sample size to investigate the extent to which it departs from its theoretical sampling distribution.

2.5 The problems under investigation

From the main theoretical background of maximum likelihood estimation in structural equation models which has been treated in the previous part of this chapter, it can be understood that the important assumptions are: independent observations, large samples and multivariate normality of the variables under study. Although it might be of interest to investigate the robustness of LISREL against dependent observations we did not do so because in the social sciences one seems to be least worried about this assumption. In many studies where observations are made on individuals it is unlikely that there are dependencies between the measurements of different persons. The reader is referred to Shaw (1978) for an example of a robustness study against dependent observations in factor analysis models.

The first goal of this research is to see how robust the LISREL-procedure is against *small sample size*. So, for a variety of models samples of specified small size are taken from multivariate normal distributions, resulting in sample covariance matrices $\underset{\sim}{S}$ on which a LISREL-analysis is performed. By repeating this a number of times the empirical sampling distributions of parameter estimates and the goodness of fit statistic can then be compared with the theoretical distributions. Thus, in the first part of this study small sample behavior is not mixed up with non-normal behavior. It depends on the small sample results whether the robustness against a combination of small samples and non-normality is worth being studied.

If the results for the small sample case are excellent or moderately good the second goal of our work will be to study the robustness of LISREL against *non-normality*. If, on the other hand, those results are bad, even for large samples, then there seems to be hardly a reason why the non-normal case is of practical interest any longer. It cannot be expected that the outcomes would be better if besides the unsufficiency of large samples an extra deficiency in assumptions is introduced. So, a full outline of the varieties in this Monte Carlo study cannot be given yet, because some of the next steps in the total research design depend on the results of the previous steps.